

Chapter 10: Matching Incoming Records to Database

In SEER*DMS, matching is a combination of automated and manual processes which determine whether a set of data is for the same patient as other data in the database. Matching algorithms are applied to several system functions, including: determining whether an incoming record contains data for an existing patient set; implementing searches in the patient lookup; and, in some registry configurations, determining whether a record is an algorithmic duplicate based on selected data fields.

This chapter describes the matching processes which are used to determine whether an incoming record contains data related to an existing patient set (linkage matches). Please see *Chapter 20: Searching for Records and Patients* and the *SEER*DMS Technical Reference* for information related to other uses of matching algorithms.

In this chapter, you'll learn about

- Automatic Matching
- Manual Review of Match Results
- Opening a Match-Consolidate Task
- Evaluating the Possible Matches
- Modifying the Reportability Status of the Incoming Record
- Requesting Follow-back Information
- Supplemental Matching Task

Automatic Matching

"Automatic matching" refers to the execution of the registry-specific algorithms that compare the patient information on an incoming record to data that were previously loaded into the database. Typically, several automated linkage matching algorithms are defined for a registry. These algorithms are associated with the various types of records loaded into SEER*DMS. The matching algorithms are implemented in XML files which can be viewed within SEER*DMS by users with the *system_administration* permission (see *Chapter 27: System Administration*).

Records that follow the passive follow-up branch of the workflow are auto-matched using registry-defined algorithms specific to passive follow-up. For these records, a Supplemental Matching Task is created if two or more patient sets are identified as "perfect" matches; or there are no perfect matches but at least one patient set is a possible match. The fields and criteria used to define a "perfect" and "possible" match are specified by the registry management. Some registries do not define "possible" matches for data used only for passive follow-up.

For each record processed in the main branch of the workflow, an automatic matching task attempts to identify patient sets and unlinked records that match the incoming record (the incoming record is not matched against unlinked records that have not yet been screened). For records that contain medical data but are not automatically used to create a patient set, a manual Match-Consolidate task is only created if possible matches are identified by the auto-match task. This enables a registrar to confirm or reject the possible matches. If the incoming record is an abstract record, a manual Match-Consolidate task is always created. The manual task is created regardless of the results of the automatic match to ensure that the match is refreshed prior to further processing of the abstract record.

The matching algorithms are re-executed when the Match-Consolidate task is opened. All available data are matched against, including data that were loaded subsequent to the initial match. The refreshing of the match guarantees that all appropriate data are included in consolidation; and ensures that new patient sets are not created from a single record when other data are available

for consolidation. This helps to reduce redundancy in worklist tasks and prevent the creation of duplicate patient sets.

Once a record completes the automated matching process, the next workflow task is determined by the results of the automatic match algorithm, the record's type, and reportability status. The following table shows the typical path for a record as it leaves the auto-match task in the main workflow.

Workflow Path of a Record as it Leaves the Automatic Match Task (Main workflow only, this does not apply to the automated Match FUP task)		
Match Results	Incoming Record	Next Step
Automated match task identified records or patient sets that may match incoming record.	Reportable or auditable records which are not NAACCR Modified or NAACCR Update records	A Match-Consolidate Task is created to allow a manual review of the possible matches that were identified.
The automated match task identified a single Patient Set that definitely matches the incoming modify or update record.	NAACCR Modified or NAACCR Update (regardless of reportability status)	Automated Consolidate FUP task is performed. If consolidation problems are encountered, a manual Consolidate FUP is created to enable someone to review and consolidate the data.
The automated match task did not identify a patient set that is a definite match but did identify unlinked records that may match; or 2 or more patient sets were identified as possible matches to the modify or update records.	NAACCR Modified or NAACCR Update (regardless of reportability status)	A Match-Consolidate Task is created to allow a manual review of the possible matches that were identified.
No matches were identified by the automatic matching task.	Reportable NAACCR Abstract or Health Records; NAACCR M records	A Match-Consolidate Task is always created for reportable abstract records. The match is refreshed when this task is opened. This ensures that the incoming record has been matched against all appropriate data, including data that were loaded subsequent to the initial match.
	NAACCR Update	An AFL is created
	Auditable NAACCR abstract or Health records; non-abstract records that are reportable	Exits the workflow but is retained in the database and held for matching to data loaded in the future.
N/A	Non-reportable records other than NAACCR M and NAACCR Update	Records that are screened in the main workflow and are not reportable are not processed in this automatic match task. They are diverted to the passive follow-up branch of the workflow for matching.

Manual Review of Match Results

A manual Match-Consolidate task is created when a manual review of the automatic match results is required, as described in the *Automatic Matching* section of this chapter. During this task, the user either selects matching data for consolidation or specifies that there are no matches. The record then moves forward in the workflow to a step that is determined by record type and reportability status. The following two tables describe the path of the record as it leaves the Match-Consolidate task.

When a user selects one or more matches and clicks Consolidate to end a Match-Consolidate task....		
Incoming Record	Matching Record(s)	Next Workflow Task
Reportable abstract record (NAACCR, NAACCR M, and Health Records are abstract records)	One or more records of any type; or a patient set	Consolidate
Reportable non-abstract record	One or more records that include a reportable abstract record; or a patient set	Consolidate
Auditable record of any type	One or more records that include a reportable abstract record; or a patient set	Consolidate
Reportable non-abstract record	One or more records – none of which are a reportable abstract record	Warning that there isn't enough data to create a patient set. Cannot be consolidated. If there is no available abstract record or patient set, select No Matches.
Auditable record of any type	A single record that is not a reportable abstract record	Non-Rpt Record Rescreening


When a user clicks No Matches to end a Match-Consolidate Task...	
Incoming Record	Next Workflow Task
Reportable abstract record	Auto-build a patient set. In some registry configurations, a Visual Edit Patient Set task will be created to enable a manual review of the data. In other configurations, the record will exit the workflow or enter a Resolve Patient Set Errors task, if appropriate.
Auditable record of any type	Exits the workflow but is held for future matching
Reportable Casefinding record	An AFL is auto-created; the Casefinding record exits the workflow but is held for future matching (see Chapter 21: Managing Abstracting Assignments)
Reportable Death Certificate record	The record is forward to Death Clearance processes (see Chapter 17: Death Clearance)
Reportable Record of any other type (Short Health, Death Notice, NAACCR Update)	Exits the workflow but is held for future matching

Opening a Match-Consolidate Task

Requires system permission: *match*

Open a Match-Consolidate task to review the records and patient sets that were determined to be “possible” matches by the automatic matching task.

To open a Match-Consolidate task:

1. Click a **Match-Consolidate** link in the worklist summary on the home page. To resume a task assigned to you, click the link listed in **My Tasks**. To open an unassigned task, click the link listed in the **Unassigned Tasks**.
2. To search for the task in your tasks and unassigned tasks:
 - a. Enter your user name in the **User(s)** filter.
 - b. Check the **Show Unassigned** box.
3. To search for a specific task, enter search criteria in the Worklist filters. Common filters to use when searching for a Match-Consolidate task are:
 - a. **Task ID(s)** – the complete task ID must be entered.
 - b. The following filters refer to data items in the incoming record:
 - i. **Data Type** – select a Record Type from the drop-down list. This only refers to the focus record’s type and not the type of records that are possible matches.
 - ii. **Record ID(s)** – a full or partial ID may be entered.
 - iii. **Facility** – a full or partial ID may be entered. Use the  lookup to select an ID.
 - iv. **DX Year(s)** – Year of diagnosis. Use a hyphen to select a range of years (e.g., 2002-2004). Use a comma to select separate years (e.g., 2002, 2004).
 - v. **Information** – enter full or partial search text related to fields in the Information column. In Match-Consolidate tasks, this includes record type and patient name. To match a complete phrase, enclose it in quotes (e.g., “Smith, John”). If you enter *John Smith* or *Smith, John* without quotes, tasks related to patients named either John or Smith will be returned (the list could include names such as John Smith, John Doe, and William Smith).
 - vi. **Tip:** To find Match-Consolidate tasks for patients whose names start with an M, try entering “; M” into the **Information** filter. You must include the quotes in your search text.
4. If you made changes to the filter settings, click **Apply**.
5. Click the task ID to open a Match-Consolidate task. Refer to the *Evaluating the Possible Matches* section of this chapter for further instructions.

Evaluating the Possible Matches

Requires system permission: *match*

When you open a Match-Consolidate Task, the incoming record and the results of the automatic matching task match will be displayed as shown below.

Incoming Record

ID	REC-3002154300 ⓘ	Type	NAACCR Abstract	SSN	999-99-9999	Site	C505: Lower-outer quadrant of breast
Facility	FAC-0089 : William Beaumont Hospital - Troy	DOB	01-01-1965	DX Date	01-25-2005	Histology	8500
Last Name	DOE	Sex	2 : Female	Race	01 : White	Behavior	3: Malignant Primary
First Name	JANE	VS	1 : Alive	Laterality	1: Right		
Middle Name	M						
Maiden		Alias		DOLC	07-25-2005		

[Follow-back](#)

Matches

5 items.

Matched in 11 seconds.

<input type="checkbox"/>	Score ▴	Type	ID ▴	Last Name ▴	First Name ▴	M ▴	DOB	SSN ▴	Sex ▴	Race	DOLC	VS ▴	Tasks
<input type="checkbox"/>	1000	Casefinding	REC-3002104467 ⓘ	DOE	JANE	M	01-01-1965	222-22-2222	2	99			
<input type="checkbox"/>	1000	Casefinding	REC-3002226399 ⓘ	DOE	JANE	M	01-01-1965		2	99			view
<input type="checkbox"/>	1000	NAACCR Modified	REC-3002453455 ⓘ	DOE	JANE	M	01-01-1965	999-99-9999	2	01	07-25-2005	1	view
<input type="checkbox"/>	499	Death Certificate	REC-3001185134 ⓘ	DOE	JANE	A...	10-10-1933	777-77-7777	2	01	05-01-1993	4	
<input type="checkbox"/>	499	Death Certificate	REC-3001459740 ⓘ	DOE	JANE	E...	05-27-1927	333-33-3333	2	01	10-17-1996	4	

[Consolidate](#)[No Matches](#)[Rematch](#)[Close](#)

If the automatic matching task identified possible matches, they will be displayed at the bottom of the page as shown in the example above. By default, the possible matches will be sorted to show the most likely matches first. The match score serves solely as a ranking mechanism and is based on registry-specific weighting schemes. The score is based on weighted rules which either consider a single data field or a combination of fields. Data fields that match the incoming record are displayed in bold font. The matching algorithm may involve fields that are not displayed such as maiden name and alias. If the score is high but fields are not shown in bold, you may review additional data fields by clicking the record or patient set IDs.

A brief summary of the incoming record is displayed at the top of the page. The primary fields used in the matching algorithm are displayed in the list of possible matches; these include name, social security number (SSN), date of birth (DOB), sex, race, vital status (VS), and date of last contact (DOLC). Although not displayed, additional fields may be used in the matching algorithms.

"Patient set" or the record type of the possible match is shown in the **Type** column. If the record is reportable, no symbol is displayed next to the record type. An **[N]** next to the record type indicates that the record is non-reportable. An **[A]** indicates that the record has a reportability status of auditable. An **[N/A]** indicates "not applicable", that is, the record type does not require screening (e.g., supplemental records are not screened for reportability).

To complete a Match-Consolidate task:

- When the Match-Consolidate task is opened, the automatic match is refreshed so that the record is matched against the most recent data.
- If the automatic matching task identified possible matches:
 - Compare the incoming record to the possible matches. Keep in mind that the list may include false positives as a result of missing data fields and the limitations of computerized matching techniques.
 - You may review additional data fields on the incoming record and on records and patient sets that are listed in the possible matches. Click the Information Icon ⓘ next to the record's or patient set's ID to see values of key fields. To browse all pages of the record or patient set, click the record's or patient set's ID. This is only available to users with system permissions enabling read and/or write access to the matching data (*rec_read_only*, *pat_read_only*, etc). Each system permission is described in *Chapter 26: Managing System Roles*.

- c. If data fields required to determine a match are missing or incomplete, click Follow-back to submit a request for additional information from the reporting source. (Refer to the *Requesting Follow-back Information* section of this chapter for further instructions.)
 - d. If necessary, you may edit data fields in the incoming record or in records or patient sets listed in the possible matches. If you make any changes to data involved in this task, click **Refresh** to refresh the screen and match.
3. If the automatic matching task does not identify any possible matches, or if you conclude that the incoming record does not match any of the patient sets or unlinked records in the list of possible matches, click **No Matches**. SEER*DMS will close the Match-Consolidate task and the record will continue to the next task in the workflow, as appropriate for the record type. If the record is an abstract record, a new patient set will ultimately be created. Other types of records will be processed via casefinding, death clearance, or follow-up processes.
 4. If you conclude that one or more of the possible matches are true matches:
 - a. Determine whether the data can be consolidated at this point. The data can be consolidated if the record is being consolidated into an existing patient set or there are sufficient data to create a new patient set. You should not consolidate data from multiple records into a new patient set if the records do not include a reportable abstract record. For example, if the incoming record is a casefinding record that only matches a casefinding record, do not consolidate the data at this point. You should click **No Matches**. A new patient set will be created when an abstract record is loaded that matches the casefinding record.
 - b. If the record is being consolidated into an existing patient set or there are sufficient data to create a new patient set, check the box next to each ID that matches the incoming record.
 - i. Click **Consolidate**.
 - ii. If the incoming and matching data include at least one reportable abstract record or patient set, a manual Consolidate task will be created in the workflow. If you have the *consolidate* system permission, the Consolidate task will open for your consideration. If not, an unassigned Consolidate task will be created in the worklist. For further instructions, refer to *Chapter 12: Consolidating Data*.
 - iii. If the incoming record was not a reportable record and it matched to one or more records that could not be used to build a patient set, a manual Non-Rpt Rescreening task will be created in the workflow. This task allows for the manual review of two non-reportable records that match each other but do not match any reportable records or patient set data. Verify the match and combine the data if appropriate. For further instructions, refer to *Chapter 15: Re-screening of Non-reportable Records*.


Modifying the Reportability Status of the Incoming Record

Requires system permission: *match* and *rec_edit*

If you determine that the incoming record of a Match-Consolidate task is incorrectly designated as reportable, you may manually modify the record's reportability status. Typically, a record's reportability status is set by an automatic screening task using a registry-defined algorithm. The reportability status may have been set by a SEER*DMS user in a manual screening task. The manual screening task is generated for any record that, according to your registry's policies, must be reviewed in order to ascertain reportability. If you determine that a record's reportability status

was set incorrectly, you should notify a registry manager. The registry-defined screening algorithm may need to be modified, or new manual screening procedures may need to be employed.

To modify the reportability status of the incoming record for a Match-Consolidate task:

1. Click the **ID** in the Incoming Record section to open the record editor.
2. In the navigation box on the left side of the editor, hold your mouse near the arrow  next to the record type. The record menu will be displayed. Select **Reportability**.
3. Select a **Reportability** value from the drop-down list.
4. If you have selected *Auditable* or *Non-Reportable*, enter text in the **Reason** field. This documentation is useful for the purpose of casefinding audits.
5. Save your changes.
 - a. Click **Save**.
 - b. Enter comments to document your changes.
 - c. Click **Save & Exit**. You will be returned to the Match-Consolidate task.
 - d. Click **Refresh** to refresh the screen and match. The icon displayed next to the record's type will indicate the new value for reportability status (no icon is shown if the record is reportable).
6. Complete the Match-Consolidate task as described in the *Evaluating Possible Matches* section of this chapter.

Requesting Follow-back Information

Requires system permission: *match* and *fb_initiate*

In SEER*DMS, a request for follow-back information is referred to as a "follow-back need". If you determine that additional information must be obtained from the reporting facility, you should submit a follow-back need. Your request will be maintained with the data record, allowing all users to review pending follow-back requests when reviewing the record data. If you cannot complete the Match-Consolidate task until you receive the information, you may close the Match-Consolidate task and complete it at a later time.

You will receive e-mail notification when a response to your request is processed and the follow-back need is closed. You or another staff member may update data fields based on the new information. As determined by registry policy, one staff member may be responsible for processing all follow-back responses, or the information may be given directly to the staff members who entered the follow-back needs.

SEER*DMS enables users to access patient data via two routes: 1) the data can be modified while completing a worklist task, or 2) the record or patient set may be edited directly, outside the context of a task. If you suspended a Match-Consolidate task pending the receipt of follow-back information, you must re-open and complete that task to allow the record to move forward in the worklist. You must either make changes to data fields based on the new information or verify that the appropriate changes were made. Specific instructions for submitting follow-back requests and processing the responses are provided in *Chapter 22: Follow-back*.

Supplemental Matching Task

Records that follow the passive follow-up branch of the workflow are auto-matched using registry-defined algorithms specific to passive follow-up. For these records, a Supplemental Matching Task is created if two or more patient sets are identified as “perfect” matches; or there are no perfect matches but at least one patient set is a possible match. The fields and criteria used to define a “perfect” and “possible” match are specified by the registry management. Some registries do not define “possible” matches for data used only for passive follow-up.

The supplemental matching algorithm only matches the incoming record against patient sets; unlinked records are not considered. The Supplemental Match Task uses the same controls and layout as the Match-Consolidate Task. The passive follow-up processes are described in *Chapter 16: Follow-up*.

To summarize, the Supplemental Match Task and the Match-Consolidate Task differ in three ways:

1. The Supplemental Match Task is part of the passive follow-up branch of the workflow. The incoming record will not be used to create a new CTC or Patient Set. It will be used to provide more current follow-up data. The Match-Consolidate Task is part of the main branch of the workflow designed for medical records that will be used to update or build CTCs and Patient Sets. Records are routed to one task or another based on record type and reportability status, as specified in registry-defined algorithms. Refer to the workflow diagrams provided with the *SEER*DMS Technical Reference* for more information.
2. In the Match-Consolidate Task, the incoming record is matched against unlinked records and patient sets. The only data not considered in this match are unlinked records that have not yet been screened for reportability. In the Supplemental Match Task, the incoming record is only matched against patient sets.
3. The auto-match algorithms related to these tasks are defined separately.